

DOI: 10.19783/j.cnki.pspc.220907

## 面向不平衡数据的配电网故障停电预测方法

范敏<sup>1</sup>, 杨青<sup>1</sup>, 郭祥富<sup>2</sup>, 刘昊<sup>3</sup>, 夏嘉璐<sup>1</sup>, 彭屿雯<sup>1</sup>

(1. 重庆大学自动化学院, 重庆 400044; 2. 国网河南省电力公司, 河南 郑州 450052;  
3. 国网河南省电力公司电力科学研究院, 河南 郑州 450052)

**摘要:** 配电网故障停电事件会严重影响正常的社会经济生活。因此, 迫切需要有效的配电网故障停电预测方法。采用人工智能方法分析配电网故障停电数据, 发现存在配电网故障停电次数较少和引发配电网故障停电的原因分布不均等数据不平衡情况。为了及时、准确地预测配电网故障停电情况, 从数据集质量和防止过拟合两方面入手改进故障停电预测模型。首先, 设计了基于聚类的对抗神经网络来增强数据集质量。其次, 构造了基于随机代价敏感卷积神经网络(RandomCost-CNN)的故障停电预测模型。RandomCost-CNN 预测算法中采用有放回随机抽样思想设计了损失函数的随机选择策略, 用以解决常规代价敏感过度拟合少数类(故障停电类)而使得大量多数类(正常类)被误报的问题, 既保证少数类具有较好召回率与精确度, 同时又提高了模型的泛化性能。实验证明所提方法能有效预测配电网故障停电事件发生概率, 在配电网运维管理中能够发挥较好的预警作用。

**关键词:** 故障停电预测; 不平衡数据分类; 过拟合; RandomCost-CNN

### Prediction method of power outage in a distribution network for unbalanced data

FAN Min<sup>1</sup>, YANG Qing<sup>1</sup>, GUO Xiangfu<sup>2</sup>, LIU Hao<sup>3</sup>, XIA Jialu<sup>1</sup>, PENG Yuwen<sup>1</sup>

(1. College of Automation, Chongqing University, Chongqing 400044, China; 2. State Grid Henan Electric Power Company, Zhengzhou 450052, China; 3. State Grid Henan Electric Power Company Electric Power Research Institute, Zhengzhou 450052, China)

**Abstract:** Power distribution network failure events will seriously affect the normal operation of social and economic life. Therefore, an effective method for predicting the power outage of a distribution network is necessary. The artificial intelligence method is used to analyze the power outage data of the distribution network. It is found that there are fewer times of power outages and the uneven distribution of the causes of the power outages in the distribution network. In order to predict the power outage situation of a distribution network in time and accurately, this paper proposes a modified power outage prediction model from the aspects of data set quality and overfitting prevention. First, it designs a cluster-based generative adversarial neural network to enhance the quality of the data set, then constructs a power outage prediction model based on a random cost-sensitive convolutional neural network (RandomCost-CNN). The RandomCost-CNN prediction algorithm adopts the idea of random sampling with replacement to design a random selection strategy of the loss functions, so that it can alleviate the minority class (power outage class) overfitting and avoid a large number of false positives for the majority class (normal class). It can ensure that minority classes have better recall & accuracy and improve the generalization performance of the model. Experiments show that the proposed method can effectively predict the probability of a power outage event in the distribution network. It can play a good early warning role in the operation and maintenance management of the distribution network.

This work is supported by the National Key Research and Development Program of China (No. 2020YFB2009405).

**Key words:** power outage prediction; imbalanced data classification; overfitting; RandomCost-CNN

## 0 引言

近年, 国家电网提出建设泛在电力物联网, 制

定数据中台, 实现“数智国网”<sup>[1-2]</sup>, 构建电力大数据创新发展生态圈, 对进一步优化管理措施, 提升服务水平具有重要意义。10 kV 配电网与用户紧密相连, 其状态将直接影响用户的用电质量。配电网接线方式多样、机理复杂、元件类型众多, 且易受

基金项目: 国家重点研发计划项目资助(2020YFB2009405)

环境因素、人为因素、运行因素等多种因素影响,故障时有发生。但线路长度较长,元件数量多,使得运维工作量大,效率低<sup>[3]</sup>。

依托于数据中台,人工智能技术可对电网建设、运维工作提供智能化辅助策略。对配电网线路故障停电风险进行预测,可及时准确地掌握配电网状态,为电网运维人员找出薄弱环节提供辅助决策支撑,有助于提高运维精细化管理水平,对保证电网可靠性、保障用户用电质量有着重要意义<sup>[4]</sup>。

目前研究中,国内外学者首先对配电网重要元件故障进行诊断与预测,提前运维检修,减少停电事件。文献[5]对配电变压器故障进行诊断,考虑到配电变压器故障记录较少的问题,使用无监督的聚类算法对配电变压器中监测的数据进行归纳分析,提取出有效的故障类别。文献[6]应用贝叶斯网络对智能电能表故障类型进行分类与预测。文献[7]建立了径向基函数(radial basis function, RBF)神经网络来分析电缆终端局部放电的问题,便于提前发现隐患。同时,许多学者考虑天气因素对配电网故障停电的影响,进行相应建模预测<sup>[8-11]</sup>。文献[8]建立了台风灾害下配电网用户停电量预测模型,将气象因素,地理因素和电网因素等作为影响因素,使用随机森林算法预测配电网用户停电量。文献[9]综合考虑故障发生频率、停电时长和缺供电量比例,确定风险等级,在此基础上提出一种计及天气因素的配电网故障特征选择和故障停电风险等级预测的方法。文献[10]使用脆弱性函数对自然灾害导致的停电事件进行预测,达到了比较好的实验效果。上述研究从多源数据中提取的特征不够丰富,故模型的泛化推广性能还有待进一步提升。

有效的特征提取方法将有助于预测模型取得更好的预测精度。文献[12]综合考虑了有源配电网的故障原因,研究故障和环境因素的关联关系,提出CNN-MSVM组合算法进行有源配电网故障等级预测。文献[13]挖掘配电网多源异构数据的关联性,提取与分析多源故障数据和辅助信息,通过逻辑规则判断配电网运行状态,诊断配电网系统中事件的属性,实现故障信息的多元挖掘与诊断。文献[14-17]通过深度学习方法尤其是卷积神经网络进行特征提取与挖掘,对配电网元器件进行故障诊断与预测,取得了较好的实验结果。由此可见,卷积神经网络具有优良的特征提取能力,是目前故障诊断与预测研究中的一个热点,因此本文以卷积神经网络算法为基础进行改进,开展故障停电分类预测研究。

从分析数据集角度来看,配电网故障停电信息相对于正常信息来说,数量较少,整个故障停电数

据集呈现类别不平衡的特点。而机器学习算法设计大多是基于均衡的数据集,在处理不平衡问题时容易将少数类误分为多数类,给建立配电网故障停电诊断与预测模型带来了一定的难度。学者们常使用代价敏感算法来解决数据不平衡问题,其关注点在于如何提高少数类的识别率,但由于加大了少数类误分类的代价,模型会过度拟合少数类,将大量多数类样本误分为少数类,造成大量误报,在实际生产环境中可能会增加不必要的工作量。除此之外,引发配电网故障停电的因素也存在不平衡的状态,在不平衡数据中还存在类内不平衡问题。学者们针对配电网故障停电数据不平衡的问题提出了相应的解决方案,如过采样<sup>[18]</sup>、进行特征筛选<sup>[19-21]</sup>、基于生成对抗神经网络(generative adversarial network, GAN)生成数据<sup>[22]</sup>等。特征筛选会损失一部分信息,过采样或生成故障数据时没有考虑配电网数据存在的类内不平衡问题,因此,模型性能还有进一步提升的空间。

针对上述问题,本文从数据质量和过拟合方面进行改进,提出一种基于随机代价敏感的卷积神经网络配电网故障停电预测模型(RandomCost-CNN)。首先对故障停电数据集进行处理,降低数据集的不平衡程度并缓解类内不平衡问题。在平衡数据的基础上设计改进的卷积神经网络模型来进一步提高分类预测性能。考虑到常规代价敏感神经网络模型过拟合少数类的问题,设计具有两个损失函数的卷积神经网络模型,并根据数据集的不平衡度设计迭代训练时损失函数的随机选择规则,缓解模型过拟合问题,最终达到提高模型预测性能的目标,实现精准预警。

## 1 配电网数据不平衡处理

### 1.1 不平衡数据集的特点

在配电网的实际生产环境中,故障停电事件是偶发事件。涉及用于故障停电预测的数据主要包括外部影响因素(如天气等)数据、设备自身影响数据和故障数据。而采集到的这些数据往往是不平衡的,具体表现为正常数据与故障数据之间实例数量相差较大。在二分类任务中常用指标 $I_R$ 来量化数据集不平衡程度,其定义如式(1)所示。

$$I_R = \frac{N_{\text{majority}}}{N_{\text{minority}}} \quad (1)$$

式中: $N_{\text{majority}}$ 为多数类(正常类)的样本个数; $N_{\text{minority}}$ 为少数类(故障停电类)的样本个数。 $I_R$ 越大则表示数据集不平衡程度越大。除此之外,不平衡数据集

中还存在类内不平衡的情况，也是影响模型分类结果的关键因素。图 1 展示了不平衡数据集存在的一个类内不平衡状态。在少数类中，样本分布也不均衡，少数类中某些子集数据量太少，如图 1 中的“少数类\_2”，模型更难学习到该类数据利于分类的特征，容易使得该类数据被大量误分为多数类。另外，有的数据偏离样本空间太远，如图 1 中的“少数类\_4”，此类噪声样本的存在可能会影响模型的分类规则，从而降低了模型对少数类的分类效果。文献[23]对某地区配电网故障数据进行统计，发现运行因素引发的故障相对于其他因素次数较少，进一步证明配电网故障数据中也存在类内不平衡问题。

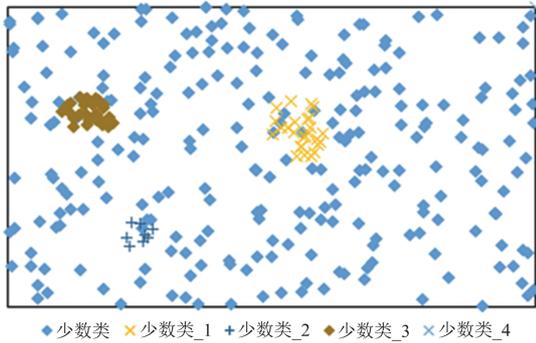


图 1 类内不平衡样本分布示例图

Fig. 1 Example of imbalanced samples with in-class

### 1.2 数据不平衡处理

由于引起故障停电事件的因素不同，故障停电数据集内部也存在样本数量不平衡的细分子类。而不平衡数据集的类内不平衡问题也是影响分类性能的一大重要因素。目前针对这方面的研究较少，基于此，本文设计基于聚类的生成对抗神经网络数据生成方法(K-means-GAN)对数据不平衡进行处理。算法示意图如图 2 所示。

步骤 1: 使用 K 均值聚类算法(K-means clustering algorithm, K-means)对少数类(故障停电类)数据集进行聚类分析，通过实验对比误差平方和残差平方和 (error sum of squares, SSE) 的值  $S_{SE}$  来选择最佳分类簇数  $K$ ，以获得相似子类。当  $S_{SE}$  不发生变化或达到最大迭代次数时，得到最终的聚类结果。 $S_{SE}$  计算公式为

$$S_{SE} = \sum_{i=1}^k \sum_{x \in C_i} |d(x, C_i)|^2 \quad (2)$$

式中:  $k$  为 K-means 聚类设置的分类簇数;  $x$  为样本实例;  $C_i$  为第  $i$  个聚类中心;  $d(x, C_i)$  为样本到聚类中心的欧氏距离。

步骤 2: 分析聚类结果，确定对少数类进行数据生成或删除的策略。若存在少数子类数量极端少，仅为个位数，且与多数类样本数量比例小于  $1:10^3$ ，将此类数据剔除。从整体上考虑若生成过多的数据，少数类数据会大量重叠；同时，也会加重数据集边界点模糊、类重叠的情况。所以，本文对少数类生成数据的数量设置了限制条件。根据聚类结果以最多的少数类子类  $D_{i_j}$  数量  $N_{i_j}$  为基准，若数据集多数类实例数量  $N$  超过 1000，则  $N/N_{i_j} > 20$  才将  $D_{i_j}$  进行成倍扩充，使  $N/N_{i_j} \leq 20$ ；若数据集多数类实例数量  $N$  不超过 1000， $N/N_{i_j} > 10$  才将  $D_{i_j}$  进行成倍扩充，使  $N/N_{i_j} \leq 10$ 。整个样本数据集越大需要生成的少数类数量就越多，则需动态调整上述阈值，避免过多生成少数类样本。

步骤 3: 利用生成对抗神经网络(GAN)模型对少数类各子类进行数据生成，缩小各少数类子类数据量大小的差距。

GAN 模型基本框架中包括生成器(Generator,  $G$ )和判别器(Discriminator,  $D$ )。对  $D$  和  $G$  的训练是一

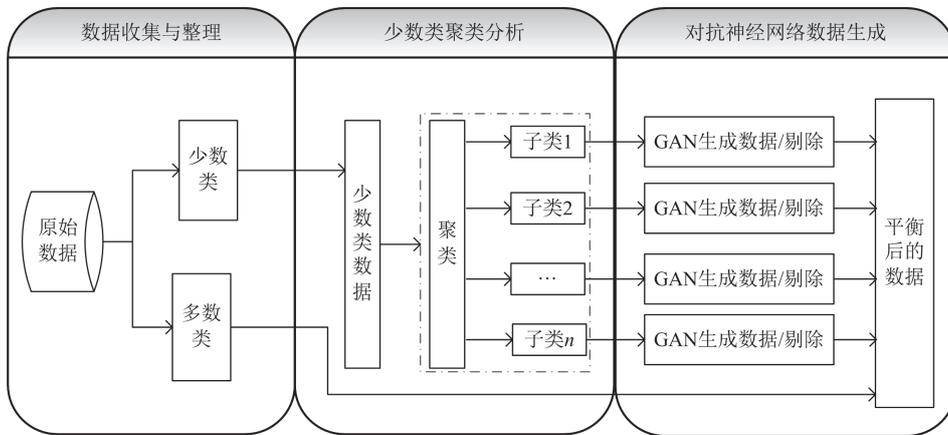


图 2 基于聚类的生成对抗神经网络数据生成示意图

Fig. 2 Schematic diagram of K-means-GAN

个极大极小的对抗过程, 模型的目标函数为

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{real}}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (3)$$

式中:  $D(x)$  表示判别器网络;  $G(z)$  表示生成器网络;  $E(\cdot)$  表示期望值;  $P_{\text{real}}(x)$  表示真实数据样本的分布;  $P_z(z)$  表示  $G$  生成的数据样本的分布。  $D$  希望式(3)最大化,  $G$  则希望其最小化, 从而形成了一个极大极小的博弈关系。求解过程分解如下所述。

对于判别器  $D$ , 其目标函数如式(4)所示。

$$\max_D V(D, G) = -(1 - y)[\log(1 - D(G(z)))] + y \log D(x) \quad (4)$$

对于生成器  $G$ , 其损失函数如式(5)所示。

$$\min_G V(D, G) = (1 - y) \log(1 - D(G(z))) \quad (5)$$

步骤 4: 将生成的各少数子类数据集与多数类数据集合并, 得到平衡后的数据集。

通过上述处理, 算法在解决少数类内部分布不平衡的同时也降低了数据集整体的不平衡程度。

## 2 随机代价敏感卷积神经网络分类预测算法

### 2.1 代价敏感

代价敏感是目前解决不平衡数据分类问题的主要手段之一, 文献[24]指出了数据挖掘中存在许多不同类型的代价, 其中之一就是误分类代价。目前常用的是基于类别的代价, 其核心思想为给不同的分类情况设置不同的分类代价, 通常使用代价矩阵表示。以二分类为例, 其代价敏感矩阵定义如表 1 所示。

表 1 代价敏感矩阵

Table 1 Matrix of cost-sensitive

类别	1(预测)	0(预测)
1(实际)	$C_{11}$	$C_{10}$
0(实际)	$C_{01}$	$C_{00}$

$C_{ij}$  表示将  $i$  类样本分类为  $j$  类样本的代价,  $i, j = 0, 1$ , 例如  $C_{10}$  表示将类别为 1 误分为类别为 0 的代价,  $C_{11}$  表示将类别为 1 正确地分类为类别 1。根据各类别的属性给不同的误分类情况设置

不同的代价。通常将正确分类的代价设置为 0。为了准确识别出少数类, 一般将少数类误分类代价设置为最大。利用代价敏感矩阵, 改变模型的优化目标, 使得期望代价最小, 可将代价敏感分类问题转化为一般的分类问题, 其优化目标如式(6)所示。

$$\arg \min \sum_{i=1}^n l_{\text{oss } i} C_{ij} \quad (6)$$

式中,  $l_{\text{oss } i}$  表示  $i$  类别的经验风险值。

卷积神经网络(convolutional neural networks, CNN)是一种前馈神经网络。它通过构建具有很多隐藏层(卷积层、池化层、全连接层等)的神经网络模型来学习更有用的特征, 从而最终提升分类的精确率。在 CNN 中常采用交叉熵损失函数作为目标函数, 在其中引入代价敏感思想, 可以学习到有利于少数类分类的特征, 从而提高少数类分类的识别率。因此, 交叉熵损失函数改进为带权重的代价敏感损失函数, 其表达式如式(7)所示。

$$l_{\text{oss}} = -[w_0 y \log \hat{y} + w_1 (1 - y) \log(1 - \hat{y})] \quad (7)$$

式中:  $y$  代表真实值;  $\hat{y}$  代表预测值;  $w_0$  和  $w_1$  分别代表了当前数据的实际标签为 1 和 0 被误分为 0 和 1 时的代价敏感系数。代价敏感系数的设置需结合其数据集的不平衡程度、数据集大小、模型训练时的表现效果等因素确定。

### 2.2 随机代价敏感卷积神经网络分类预测模型

常规的代价敏感卷积神经网络模型着眼于追求少数类的高召回率, 会使模型过度拟合少数类, 从而使得大量的多数类样本被误分为少数类, 降低了模型的性能[25]。在遗传算法中, 为保证样本多样性, 并不是每次都选择适应度最高的遗传算子, 而是按照一定的策略, 如轮盘赌、锦标赛选择法等策略, 随机从父代中挑选一些个体生存下来, 适应度高的遗传算子被选择的概率较大, 既保证了从群体中选择优胜的个体又保证了样本多样性, 缓解出现局部最优解的问题。本文借鉴上述思路, 提出了随机代价敏感卷积神经网络(RandomCost-CNN)分类预测模型, 其基本结构如图 3 所示。将一般的损

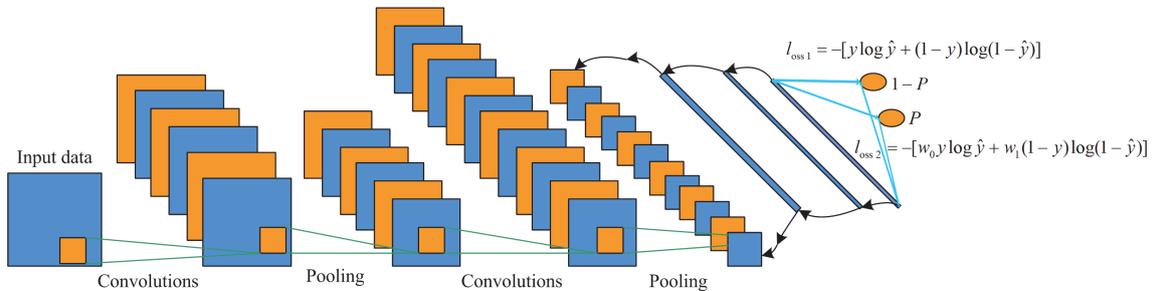


图 3 RandomCost-CNN 基本结构

Fig. 3 Structure of RandomCost-CNN model

失函数  $l_{oss1}$  (式(8))与代价敏感损失函数  $l_{oss2}$  (式(9))共同作为卷积神经网络分类模型的损失函数,并在每次迭代训练时随机选择其中之一作为此次训练的损失函数,从而防止模型过拟合。

$$l_{oss1} = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (8)$$

$$l_{oss2} = -[w_0 y \log \hat{y} + w_1 (1 - y) \log(1 - \hat{y})] \quad (9)$$

为方便调整代价敏感系数,本文将  $w_1$  设置为 1,只设置少数类误分的代价敏感系数  $w_0$ ,其初始值为数据集的不平衡度  $I_R$ ,后续需根据模型训练结果进行优化,调整  $w_0$  的值。

根据数据集不平衡程度制定相应的随机选择机制。其选择算法借鉴有放回随机抽样法思想:即随机从总体中抽取样本,并将抽取的样本转换为概率。假设概率之和为 1 即为全部样本,从 [0.00,1.00] 中随机抽取一个小数  $P_i$ ,  $P_i$  为精确到百分位的小数,若  $P_i$  小于设定的阈值  $P$ ,则选择  $l_{oss2}$  为该次训练的损失函数,否则选择  $l_{oss1}$  为该次迭代训练的损失函数。其中  $P$  的初始值为式(10),后续需根据分类效果对模型优化,调整  $P$  的值。

$$P = \frac{I_R}{10} \quad (10)$$

一般地,神经网络训练次数为几千到几万次,根据大数定律(式(11)),在随机大量重复的事件中,往往呈现出一定的规律。所以,损失函数的选择具有一定的随机性的同时也基本符合设置的规则。

$$\lim_{n \rightarrow \infty} P \left( \left| \frac{\mu_n}{n} - p \right| < \varepsilon \right) = 1 \quad (11)$$

式(11)中,假设事件  $A$  表示在单次模型训练中选择代价敏感损失函数作为该次训练损失函数的行为,其发生的概率为  $p$ ,  $\mu_n$  为  $n$  次独立随机选择中  $A$  发生的次数,  $\varepsilon$  为任意正整数,则事件  $A$  出现的概率基本为其发生的概率。

### 2.3 算法设计

RandomCost-CNN 分类预测算法设计如下。

步骤 1: 输入经过一定平衡后的数据集  $D_{atase}$ , 计算数据集的  $I_R$ 。

步骤 2: 根据数据集的不平衡程度设计相应的代价敏感损失函数  $l_{oss1}$  和  $l_{oss2}$ , 并设置  $l_{oss2}$  的  $w_0$  为  $I_R$ ,  $w_1$  为 1.0。

步骤 3: 划分训练集和验证集,搭建卷积神经网络训练模型。设置模型的训练次数  $i$  和每轮训练次数 ( $j = N / n_{batch\_size}$ ), 其中,  $n_{batch\_size}$  为每次模型训练所抽取的数据样本容量。在每轮次训练(第  $j$  轮)中,从 [0.00,1.00] 中随机抽取一个小数  $P_i$ , 若  $P_i$  小

于概率  $P$ , 则选择损失函数为  $l_{oss2}$ , 否则选择损失函数为  $l_{oss1}$ ; 以最小化损失函数为目标训练卷积神经网络模型。经训练迭代后得到分类预测模型。

步骤 4: 优化模型,优化  $w_0$ 、 $P$ 、训练次数等参数。

### 3 配电网故障停电预测算法

配电网故障停电预测属于典型的分类预测问题。故障停电事件在配电网的实际生产环境中属于偶发事件。在采集配电网馈线数据时,往往是正常数据量远大于故障数据量。因此,解决数据集的不平衡程度问题是故障停电预测分类模型提升预测精度的难点之一。同时,由于引起配电网馈线故障停电事件的影响因素众多,包括天气、时间、季节、运行情况、设备数据等,如何有效提取影响配电网故障停电的特征对于建立高质量的故障停电预测模型至关重要。卷积神经网络在提取非线性特征方面具有突出的优势。本文设计的随机代价敏感卷积神经网络(RandomCost-CNN)分类预测模型不但具有特征提取的优势,同时兼顾不平衡数据集处理与防止少数类过拟合的特点,适用于面向配电网故障停电预测问题建立较精准的预测模型。通过模型预测可以实现配电网馈线故障停电预警,提前进行隐患消除,有助于保障配电网的供电可靠性,提高运维工作效率。

本文基于 RandomCost-CNN 的配电网故障停电预测算法研究主要分为 3 个部分: 1) 配电网故障停电数据前期准备工作; 2) 利用 K-means-GAN 方法对配电网故障停电数据不平衡处理; 3) 配电网故障停电预测模型构建。具体流程如图 4 所示。

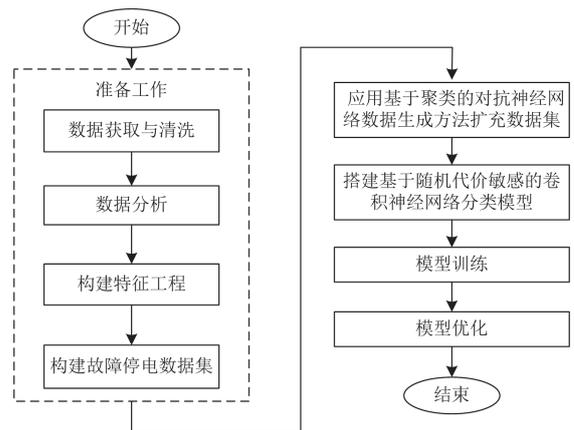


图 4 配电网故障停电预测流程图

Fig. 4 Flowchart of distribution network fault outage prediction

步骤 1: 准备工作。根据任务获取电力系统中与任务相关的数据,通过爬虫获取外部天气数据等。

对原始数据做预处理工作, 构建故障停电数据集, 为后续分析建模做准备。

步骤 2: 配电网故障停电存在数据不平衡及类内不平衡问题, 运用 K-means-GAN 方法对数据不平衡进行处理。

步骤 3: 模型训练与优化。将预处理后的数据集按一定规则划分为训练集和验证集, 运用 RandomCost-CNN 分类预测算法构建配电网馈线故障停电预测模型。利用训练集数据训练模型, 根据模型预测结果调节参数。利用验证集测试模型, 验证模型的预测精度、召回率等, 判断模型是否达到预期目标, 并在此基础上进行优化。

## 4 实验分析

### 4.1 数据获取

配电网大数据可分为两部分, 一部分是配电网内部数据, 另一部分是利用爬虫技术获取的天气数据, 根据需求获取表 2 所示的数据。

表 2 预测所需数据

类别	来源	类型
外部影响因素数据	中国天气网等	温度、风速、湿度、降雨量等
设备自身影响数据	SCADA 等	线路基本信息、所属变电站、隐患记录等
故障数据	PMS 等	故障线路、时间、类型、停电台区数、故障原因等

### 4.2 特征工程提取

本文对获取的 2013—2017 年某地区配电网馈线数据进行分析, 以 59 条发生过故障停电事件的馈线为对象, 故障停电预测的时间尺度设置为周, 从设备因素、气象因素、故障自相关性等方面进行特征提取与构造(共计 47 个特征), 建立故障停电数据集。整个数据集共有 15 340 条数据, 其中, 正常的数据 14 415 条, 发生停电的数据 925 条, 正负样本间数量差距较大, 不平衡度  $I_R = 15.58$ , 属于典型的不平衡数据集。

### 4.3 不平衡数据集处理

将 925 个少数类样本全部提取出来利用 K-means 算法进行聚类。选取  $K$  为 2 到 7, 其评价指标 SSE 的变化如图 5 所示。根据聚类结果, 当  $K=3$  时, 首次出现 SSE 变化的最大转折点, 故确定聚类的簇为 3, 得到图 6 的聚类结果。每个少数类的数量如表 3 所示, 由此可见少数类内部也存在数据不平衡。

分析少数类聚类结果, 确定生成策略。数据集中多数类数量  $N$  为 14 415, 最大少数类子类 Class 2

为 619, 则根据生成规则  $14115/619 > 20$ , 因此本文先将 Class 2 扩充为原来的 2 倍, 然后以扩充后的 Class 2 为基准对其他子类数据进行扩充, 使各子类样本数量呈倍数增加后与 Class 2 样本数量相差不大。

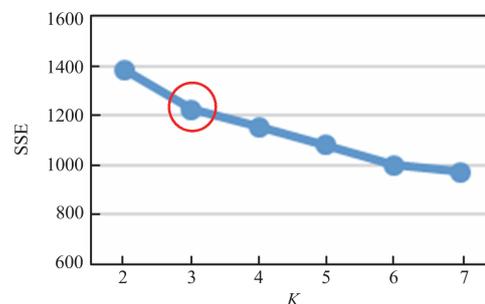


图 5 聚类 SSE 变化图

Fig. 5 Change chart of clustering SSE

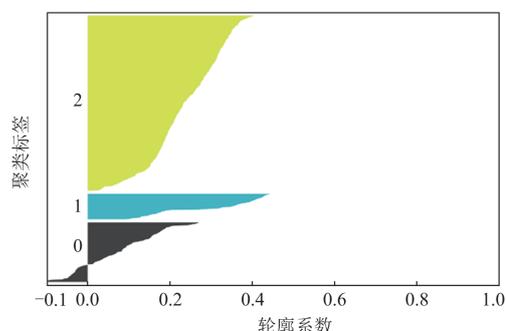


图 6 聚类结果

Fig. 6 Result of clustering

表 3 聚类子样本数量

Class	数量
Class 0	211
Class 1	95
Class 2	619

搭建 GAN 模型生成数据, 因配电网相关数据各属性间数据值相差较大, 如属性绝缘化率范围为  $[0,1]$ , 而属性线路总长度可达上百千米, 因此先将数据进行最大最小归一化, 再进行数据生成。另外, 本文所涉及到的数据维数不高, 所以使用常用的反向传播(back propagation, BP)神经网络作为生成器生成数据, 其网络为 3 层, 模型结构如图 7 所示。生成器  $G$  中的输入数据采用服从正态分布的  $47 \times 1$  的随机数据, 经过 3 层 BP 神经网络后输出生成数据, 每层都使用 ReLU 作为激活函数。判别器  $D$  判别是真实数据还是模拟数据, 其网络结构也是采用 3 层 BP 神经网络, 最后一层采用 sigmoid 激活函数用于分类, 其余两层使用 ReLU 作为激活函数。GAN

模型训练采用训练一次判别器  $D$  再训练一次生成器  $G$  的方式。其中,  $D$  使用的学习率为 0.000 15,  $G$  使用的学习率为 0.0001。确定最终迭代次数为 3000, 每个类别选取其中一组数据进行分析, 数据生成效果如图 8 所示。

图 8 中横坐标表示数据集的属性, 纵坐标表示属性的值; 折线中蓝色代表原始数据, 绿色代表模型生成的数据, 可见每个子类间数据分布存在一定的差异, 符合聚类后的结果, 其生成数据走势基本与原始数据吻合。少数类扩充后数据集样本数量如表 4 所示, 表明经预处理后数据集的不平衡度在一定程度上减小了。

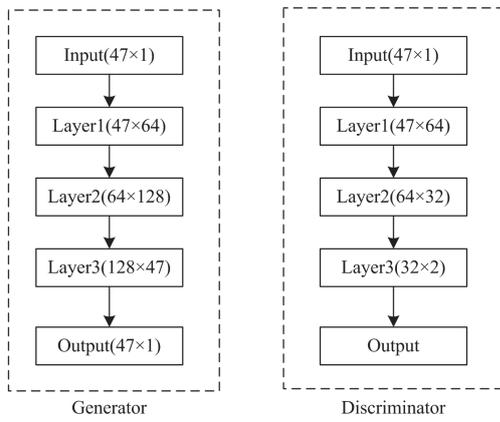


图 7 GAN 模型结构

Fig. 7 Structure of GAN model

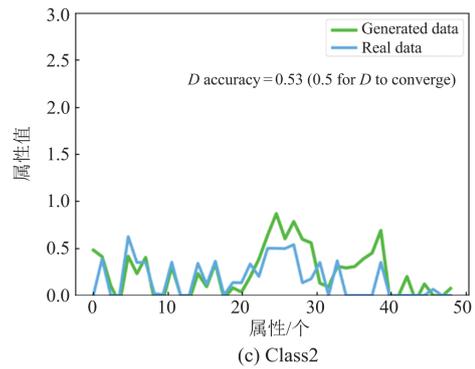
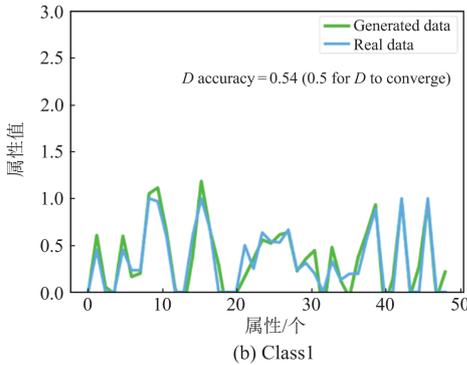
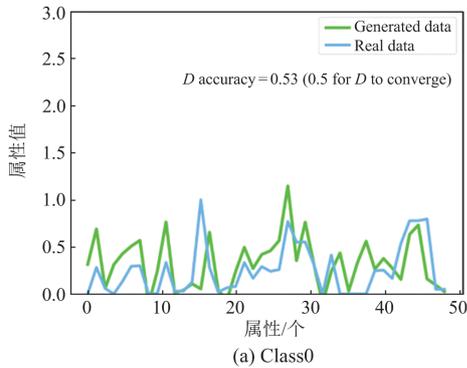


图 8 生成数据分析

Fig. 8 Analysis of generated data

#### 4.4 配电网故障停电预测模型构建

##### 1) 设计损失函数

模型设计了两个损失函数, 如式(8)和式(9)所示。根据数据集的  $I_R$  确定代价敏感系数  $w_0$  的初始值为 3.86。

表 4 数据集整体情况

Table 4 Overall situation of the data set

数据集	多数类:少数类	不平衡度
故障停电数据集	14415:3739	3.86

##### 2) 构造卷积特征

卷积神经网络对应的输入为二维的数据, 而本文构造的故障停电数据集是一维的, 所以需将一维的数据集转换为二维数据集。本文提取的特征有 47 个, 在此基础上填充两个 0 变换为 49 个特征, 将其转换为  $7 \times 7$  的二维数据集, 其转换示意图如图 9 所示。

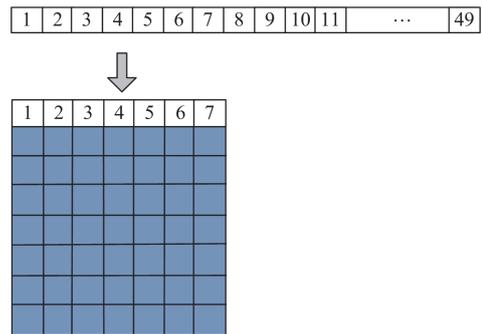


图 9 数据转换示意图

Fig. 9 Chart of data conversion

##### 3) 设计 RandomCost-CNN 分类模型

本文所搭建的 RandomCost-CNN 分类模型结构如图 10 所示, 输出层对应两个损失函数, 利用神经网络分批次多轮训练的特性, 在训练过程中随机选择其中一个损失函数作为该轮训练的损失函数, 每次训练时损失函数是不确定的。简单来说, 代价敏

感算法提高了少数类误分类代价, 并以全局最小误分代价为最小化目标, 在大多数情况下会过度拟合少数类。为了保证模型的泛化性, 本文不仅使用具有代价敏感的损失函数作为最小化目标来更新模型连接的权重, 也在模型训练中随机采用一般的损失函数作为最小化目标来更新模型连接的权重, 防止模型过度拟合少数类, 以提高模型的泛化性。

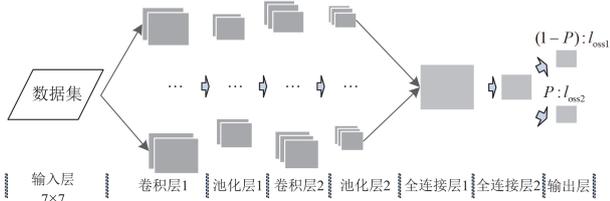


图 10 RandomCost-CNN 模型结构

Fig. 10 Structure of RandomCost-CNN model

#### 4) 模型训练

实验使用 5 折交叉验证对模型进行训练及优化。在本次实验中, 不平衡度  $I_R$  为 3.86, 则初始化损失函数选择的阈值  $P$  为 0.38。根据验证集实验结果对模型进行优化, 调整模型代价敏感系数、随机选择代价敏感函数的阈值  $P$ 、模型训练次数等参数, 选择分类性能较好的参数组合, 最终参数如表 5 所示。

表 5 模型参数组合

Table 5 Combination parameter of model

数据集	训练轮数	$w_0$	$P$
故障停电数据集	6500	9.0	0.85

#### 4.5 实验结果分析

本文基于实际生产环境中的配电网相关数据构造数据集(以下实验没有明确说明的, 都是采用经 K-means-GAN 方法处理后的数据集), 按比例划分训练集与测试集, 并通过评价指标预测召回率和 F1 衡量模型的分类效果。

##### 1) 实验一

本实验对比分析了代价敏感卷积神经网络(loss-function cost-sensitive convolutional neural network, LC-CNN)和本文所提 RandomCost-CNN 方法的预测分类效果。RandomCost-CNN 参数设置如表 5 所示, LC-CNN 训练次数及代价敏感系数同表 5, 其预测分类混淆矩阵如图 11 所示。

从图 11 可知, LC-CNN 提高了少数类的识别率, 未预测出的少数类个数由 61 降到了 58; 但多数类误判为少数类的个数大大提高了, 由原来的 40 增加到了 123, 代价敏感分类算法提高了少数类的识别率, 但会使得模型过度拟合少数类, 将多数类

样本误分为少数类, 造成大量误报。LC-CNN 及 RandomCost-CNN 的召回率及 F1 的预测分类结果如表 6 所示。

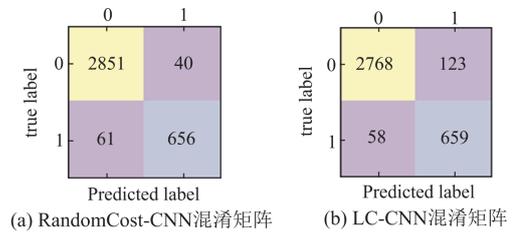


图 11 两种模型的预测混淆矩阵

Fig. 11 Predictive confusion matrix of two models

表 6 两种模型的预测结果

Table 6 Predictive effects of two models

模型	召回率	F1
LC-CNN	<b>0.919</b>	0.879
RandomCost-CNN	0.915	<b>0.929</b>

从表 6 可知, LC-CNN 提高了 0.4% 的少数类识别率, 但是忽略了少数类的精确率, 其 F1 大幅降低 5%, 使得大量多数类误分为少数类, 本文所提算法保证了模型具有较高的召回率的同时也具有较高的精确率, 其预测性能较好。

##### 2) 实验二

本实验对比分析了不平衡数据处理方法 K-means-GAN 的应用效果。对比实验设置如下: (1) 应用 LC-CNN 和 RandomCost-CNN 在不做任何处理的数据集上进行分类; (2) 应用 LC-CNN 和 RandomCost-CNN 在搭建 K-means-GAN 模型生成数据的基础上进行分类。F1 值对比如图 12 所示。

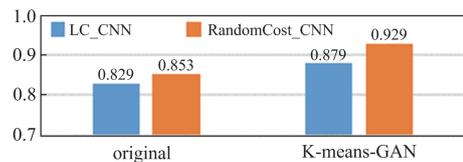


图 12 F1 对比图

Fig. 12 Comparison chart of F1

从图 12 可知, 使用本文所提 K-means-GAN 方法进行数据集预处理后, 再运用 LC-CNN 和 RandomCost-CNN 进行故障停电预测, 预测精度均有所提升; 并且本文提出的 RandomCost-CNN 方法预测效果更好, 其 F1 值较优。

##### 3) 实验三

为了进一步验证 RandomCost-CNN 方法的有效性, 分别同常规分类预测算法随机森林、XGBoost、

支持向量机(Support vector machine, SVM)和 BP 神经网络进行预测效果对比, 相应预测混淆矩阵如图 13 所示。

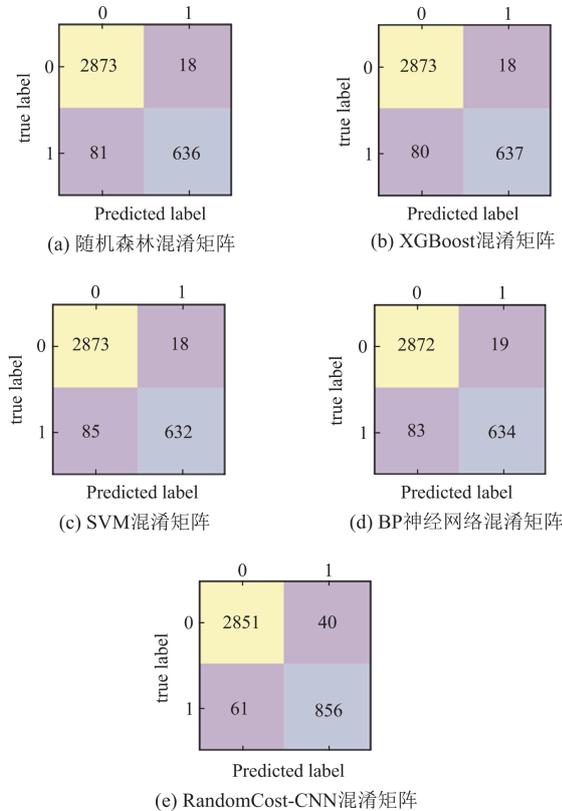


图 13 5 种模型的预测混淆矩阵

Fig. 13 Predictive confusion matrix of five models

由图 13 可知, 在对 717 个少数类故障停电样本进行预测时, 随机森林、XGBoost、BP、SVM 未识别出少数类的个数均在 80 及以上, 而本文所提的算法提高了少数类的误分类代价, 其未识别出少数类的个数降到了 61, 大大提高了少数类的识别率。5 种模型的召回率及 F1 的预测分类结果如表 7 所示。

表 7 5 种模型预测效果

Table 7 Predictive effects of five models

模型	召回率	F1
随机森林	0.887	0.928
XGBoost	0.888	<b>0.929</b>
SVM	0.881	0.923
BP	0.884	0.926
RandomCost-CNN	<b>0.915</b>	<b>0.929</b>

从表 7 可知, 随机森林、XGBoost、BP、SVM 对少数类的召回率都在 0.89 以下, 而本文所提算法 RandomCost-CNN 对少数类的预测召回率为 0.915。F1 为精确率和召回率的复合指标, 本文所提算法与 XGBoost 的 F1 值最高, 说明在提高少数类的召回

率的同时, 也保证了模型整体预测的精确率, 模型性能具有较好预测分类效果。

4) 实验四

对本文所提算法稳定性进行分析, 取 100 次实验的平均值与标准差系数进行说明, 结果如表 8 所示。表 8 中, 召回率均值为 0.916, F1 均值为 0.931, 均比实验三中的算法效果好。在统计学中常使用标准差系数来度量一组数据的离散程度。因召回率及 F1 均小于 1, 其标准差值相对较小, 所以本实验使用标准差系数  $\delta$  来衡量离散程度, 将标准差与其相应的平均数  $\bar{x}$  对比, 计算标准差系数, 采用相对数进行度量。

表 8 算法稳定性实验结果

Table 8 Algorithm stability experiment results

	召回率	F1
$\bar{x}$	0.916	0.931
$\delta$	0.0029	0.0055

计算 100 次实验召回率和 F1 的标准差系数分别为 0.0029 和 0.0055, 均小于 1%, 说明算法稳定性较好。

5) 应用效果

本文以一条实际线路数据为例来说明模型的预测效果。选取某条线路 2013 年 1 月 1 日—3 月 31 日这 3 个月的数据进行测试(预测时段以周为单位, 且该线路数据未参与模型训练), 模型测试结果如图 14 所示, 图中: 纵轴表示停电风险, 横轴对应日期; 黄色的线表示真实情况, 蓝色的线表示预测值。从图 14 可以看到, 模型预测出在第 2 周该线路发生故障停电的风险为 0.91, 表示该段时间发生故障停电的可能性极大。而该线路在 2013 年 1 月 13 日(第 2 周)确实发生故障停电事件, 说明模型的结果比较准

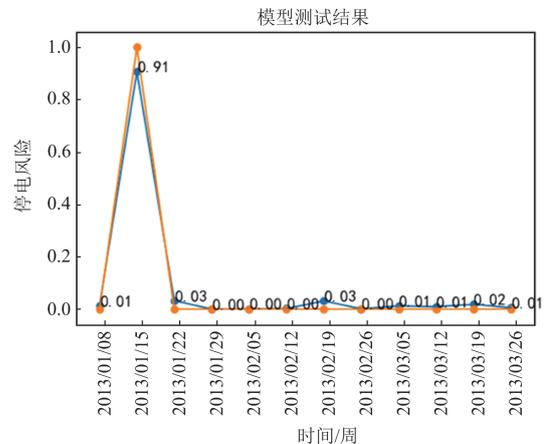


图 14 模型测试结果

Fig. 14 Test result of model

确, 具有较大的参考意义, 可为电网人员的运维检修工作提供参考, 提前进行隐患消缺, 保证电网安全运行。

## 5 结论

本文提出的配电网馈线故障停电预测方法经分析验证可以得出以下结论:

1) 应用 RandomCost-CNN 预测分类算法相对于应用一般的代价敏感分类算法, 提高了 F1 值, 缓解了模型过度拟合少数类的问题, 提高了模型预测效果, 在提高少数类的召回率的同时保证了较高的精确率;

2) 应用 K-means-GAN 方法对不平衡数据进行预处理, 可以有效改善分类预测模型的精度;

3) 横向对比了常规分类预测模型随机森林、XGBoost、BP 和 SVM 算法的预测效果, 本文所提 RandomCost-CNN 方法在提高少数类识别率的同时, 提高了模型整体的 F1 值, 其预测性能较好;

4) 对模型的稳定性进行了分析, 100 次实验结果相差不大, 其稳定性较好;

5) 本文提出了一种基于改进卷积神经网络的配电网故障停电预测模型, 但目前获取数据还有一定限制, 随着泛在电力物联网的建设, 未来采集到的配电网相关数据会更加完善、精准且粒度更小, 可进一步提升模型的预测效果, 在实际应用中可为电网人员提供更精准的决策信息。

## 参考文献

- [1] 肖振峰, 辛培哲, 刘志刚, 等. 泛在电力物联网形势下的主动配电网规划技术综述[J]. 电力系统保护与控制, 2020, 48(3): 43-48.  
XIAO Zhenfeng, XIN Peizhe, LIU Zhigang, et al. An overview of planning technology for active distribution network under the situation of ubiquitous power internet of things[J]. Power System Protection and Control, 2020, 48(3): 43-48.
- [2] CHEN Haoyang, WANG Xiaojuan, LI Zhihao, et al. Distributed sensing and cooperative estimation/detection of ubiquitous power internet of things[J]. Protection and Control of Modern Power Systems, 2019, 4(2): 151-158.
- [3] 李祖明, 孙仲明, 顾乔根, 等. 基于继电保护装置态势感知及辅助决策的智能运维系统[J]. 电力系统保护与控制, 2020, 48(19): 142-150.  
LI Zuming, SUN Zhongming, GU Qiaogen, et al. Smart operation and maintenance system based on situational awareness and assistant decision-making of relay protection devices[J]. Power System Protection and Control, 2020, 48(19): 142-150.
- [4] 曹华珍, 吴亚雄, 李浩, 等. 基于海量数据的多维度负荷特性分析系统开发[J]. 电力系统保护与控制, 2021, 49(6): 155-166.  
CAO Huazhen, WU Yaxiong, LI Hao, et al. Development of a multi-dimensional load characteristic analysis system based on massive data[J]. Power System Protection and Control, 2021, 49(6): 155-166.
- [5] DIAS L, RIBEIRO M, LEITO A, et al. An unsupervised approach for fault diagnosis of power transformers[J]. Quality and Reliability Engineering International, 2021, 37(6): 2834-2852.
- [6] 郑安刚, 张密, 曲明钰, 等. 基于贝叶斯网络的智能电表故障类型预测[J]. 电测与仪表, 2018, 55(21): 143-147.  
ZHENG Angang, ZHANG Mi, QU Mingyu, et al. Prediction of the fault type of smart meters based on the Bayesian network[J]. Electrical Measurement & Instrumentation, 2018, 55(21): 143-147.
- [7] ABDELKADER E, TAHAR S, ZEGNINI B, et al. Partial discharge activity diagnosis in electrical cable terminations using neural networks[J]. Transactions on Electrical and Electronic Materials, 2021, 22(6): 904-912.
- [8] LI M, HOU H, YU J, et al. Prediction of power outage quantity of distribution network users under typhoon disaster based on random forest and important variables[J]. Mathematical Problems in Engineering, 2021(6): 1-14.
- [9] 张稳, 盛万兴, 刘科研, 等. 计及天气因素相关性的配电网故障风险等级预测方法[J]. 电网技术, 2018, 42(8): 2391-2398.  
ZHANG Wen, SHENG Wanxing, LIU Keyan, et al. A prediction method of fault risk level for distribution network considering correlation of weather factors[J]. Power System Technology, 2018, 42(8): 2391-2398.
- [10] ZHAI C, CHEN Y J, WHITE A G, et al. Power outage prediction for natural hazards using synthetic power distribution systems[J]. Reliability Engineering System Safety, 2020, 208.
- [11] MADHIARASAN M. Accurate prediction of different forecast horizons wind speed using a recursive radial basis function neural network [J]. Protection and Control of Modern Power Systems, 2020, 5(3): 230-238.
- [12] 岳震. 基于深度学习的有源配电网故障定位与预测研究[D]. 北京: 华北电力大学, 2021.  
YUE Zhen. Research on faulty location and prediction of active distribution network based on deep learning[D]. Beijing: North China Electric Power University, 2021.
- [13] 马洲俊, 樊飞龙, 王勇, 等. 基于多源异构数据的配电网故障信息挖掘与诊断[J]. 供用电, 2018, 35(8): 31-39.

- MA Zhoujun, FAN Feilong, WANG Yong, et al. Distribution network fault information mining and diagnosis based on multi-source heterogeneous data[J]. *Distribution & Utilization*, 2018, 35(8): 31-39.
- [14] 周王峰, 李勇, 郭钊秀, 等. 基于 DAE-LSTM 神经网络的配电网日线损率预测[J]. *电力系统保护与控制*, 2021, 49(17): 48-56.
- ZHOU Wangfeng, LI Yong, GUO Yixiu, et al. Daily line loss rate forecasting of a distribution network based on DAE-LSTM[J]. *Power System Protection and Control*, 2021, 49(17): 48-56.
- [15] 李宇, 杨柳林. 基于卷积神经网络的配电网单相接地故障识别[J]. *电气工程学报*, 2020, 15(3): 22-30.
- LI Yu, YANG Liulin. Identification of single-phase-to-earth fault in distribution network based on convolutional neural network[J]. *Journal of Electrical Engineering*, 2020, 15(3): 22-30.
- [16] 常淼, 沈艳霞. 基于改进卷积神经网络的风电轴承故障诊断策略[J]. *电力系统保护与控制*, 2021, 49(6): 131-137.
- CHANG Miao, SHEN Yanxia. Fault diagnosis strategy of a wind power bearing based on an improved convolutional neural network[J]. *Power System Protection and Control*, 2021, 49(6): 131-137.
- [17] 孙曙光, 李勤, 杜太行, 等. 基于一维卷积神经网络的低压万能式断路器附件故障诊断[J]. *电工技术学报*, 2020, 35(12): 2562-2573.
- SUN Shuguang, LI Qin, DU Taihang, et al. Fault diagnosis of accessories for the low voltage conventional circuit breaker based on one-dimensional convolutional neural network[J]. *Transactions of China Electrotechnical Society*, 2020, 35(12): 2562-2573.
- [18] 杨华. 配电网故障预测的研究与应用[D]. 沈阳: 中国科学院大学(中国科学院沈阳计算技术研究所), 2020.
- YANG Hua. Research and application of distribution network fault prediction[D]. Shenyang: Shenyang Institute of Computing Technology, Chinese Academy of Science, 2020.
- [19] 杨凤生, 熊波, 蔡广林, 等. 基于数据分析的配电网故障数据特征变量提取[J]. *电力大数据*, 2018, 21(3): 22-26.
- YANG Fengsheng, XIONG Bo, CAI Guanglin, et al. Characteristic variable extraction of fault data in distribution feeder based on data analysis[J]. *Electric Power Big Data*, 2018, 21(3): 22-26.
- [20] 高小芹, 王梓玮, 卢晨, 等. 基于最小二乘支持向量机的配电网故障量预测模型[J]. *武汉理工大学学报(信息与管理工程版)*, 2020, 42(1): 23-29.
- GAO Xiaoqin, WANG Ziwei, LU Chen, et al. Prediction model of distribution network faults based on least squares support vector machine[J]. *Journal of Wuhan University of Technology (Information and Management Engineering Edition)*, 2020, 42(1): 23-29.
- [21] 彭和平, 莫文雄, 王勇, 等. 基于配电大数据的配电网故障停电影响因素灵敏度分析[J]. *电力信息与通信技术*, 2021, 19(8): 61-68.
- PENG Heping, MO Wenxiong, WANG Yong, et al. Sensitivity analysis of influencing factors of power failure in distribution network based on big data[J]. *Electric Power Information and Communication Technology*, 2021, 19(8): 61-68.
- [22] 孔洁. 基于循环条件生成对抗网络的数据生成方法[J]. *火控与指挥控制*, 2021, 46(11): 134-139, 143.
- KONG Jie. Data generation method based on cyclic conditions to generative adversarial network[J]. *Fire Control & Command Control*, 2021, 46(11): 134-139, 143.
- [23] 刘亚玲. 基于深度学习的 10 kV 馈线故障预测方法研究[D]. 重庆: 重庆大学, 2019.
- LIU Yaling. Research on 10 kV feeder fault prediction method based on deep learning[D]. Chongqing: Chongqing University, 2019.
- [24] TURNEY P. Types of cost in inductive concept learning[C]// *Proceedings of the Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning*, June 29-July 2, 2000, Stanford, USA: 15-21.
- [25] 陈琼, 徐洋洋, 陈林清. 不平衡数据的迁移学习分类算法[J]. *华南理工大学学报(自然科学版)*, 2018, 46(1): 122-130.
- CHEN Qiong, XU Yangyang, CHEN Linqing. Transfer learning for classification on unbalanced data[J]. *Journal of South China University of Technology (Natural Science Edition)*, 2018, 46(1): 122-130.

收稿日期: 2022-06-16; 修回日期: 2022-10-11

作者简介:

范敏(1975—), 女, 通信作者, 副教授, 主要研究方向为机器学习与大数据处理; E-mail: fanmin@cqu.edu.cn

杨青(1995—), 女, 硕士研究生, 主要研究方向为机器学习与大数据处理; E-mail: 506393014@qq.com

郭祥富(1983—), 男, 硕士, 教授级高工, 主要研究方向为配电网物联网、电力信息化。E-mail: 27667767@qq.com

(编辑 周金梅)